Baer Invariants and the Birkhoff–Witt Theorem

P. J. Higgins

Department of Mathematics, King's College, Strand, London W.C. 2

Received July 4, 1968

Published in: *Journal of Algebra* **11** (1969), pp. 469–482.

Digitized with Claude Opus 4.7 based on a scan from Elsevier Open Archive
(original scan: doi:10.1016/0021-8693(69)90086-6).

Proofread by Darij Grinberg 21 May 2026. Appendices added by Darij Grinberg 22 May 2026.

1. Introduction

The Birkhoff–Witt theorem asserts that, under certain conditions on a Lie algebra M (over a commutative ring R), the graded algebra associated with the enveloping algebra of M is canonically isomorphic with the symmetric algebra of the underlying module of M. It then follows that M is embedded in its enveloping algebra. It is known that not all Lie algebras are so embedded, but the theorem has been proved in the following cases:

(i) when M is a free R-module, for any R (Birkhoff [1], Witt [7]);

(ii) when R is a Dedekind domain, for any M (Lazard [5], Cartier [2]);

(iii) when R is an algebra over the rationals, for any M (Cohn [3]).

These conditions refer only to the module structure of M, and we aim to give a unified treatment in which the Lie product on M and the enveloping algebra of M are eliminated from the discussion at an early stage. They are replaced by a homological invariant $B(M)$, defined for an arbitrary module M, which depends only on the tensor algebra of M and is one of the Baer invariants defined for algebras by Fröhlich [4]. The vanishing of $B(M)$ implies the Birkhoff–Witt theorem for all Lie algebras with underlying module M, and we shall show that $B(M) = 0$ in the three cases mentioned above. In fact the condition $B(M) = 0$ is necessary and sufficient for the validity of a closely related embedding theorem for "Lie structures over M" in "associative structures over M." The definition of these concepts is our starting-point.

2. Lie Structures and Associative Structures

Let R be a commutative ring (with identity) and let M be an R-module. We denote by $T(M) = \bigoplus_{n\geq 0}(\otimes^n M)$ the tensor algebra of M and identify M with the homogeneous part of $T(M)$ of degree 1. (All tensor products are taken over R unless otherwise stated). The canonical map from $T(M)$ to the symmetric algebra $S(M)$ has as kernel the ideal of $T(M)$ generated by all commutators $xy - yx$ $(x, y \in M)$. We denote this kernel by $K(M)$; it is a homogeneous ideal of $T(M)$.

If we consider $K(M)$ as a $T(M)$-bimodule, it is generated by the commutators $xy - yx$, and these satisfy some obvious relations which we use as axioms in the following definition. A *Lie structure over the R-module M* is a $T(M)$-bimodule A together with a bilinear function $M \otimes M \to A$ (denoted by $x \otimes y \mapsto \langle x, y\rangle$), satisfying the axioms:

(L1) $\langle x, x\rangle = 0 \quad (x \in M)$;

(L2) $\langle x, y\rangle t(uv - vu) = (xy - yx)t\langle u, v\rangle \quad (x, y, u, v \in M,\ t \in T(M))$;

(L3) $(\langle x, y\rangle z - z\langle x, y\rangle) + (\langle y, z\rangle x - x\langle y, z\rangle) + (\langle z, x\rangle y - y\langle z, x\rangle) = 0 \quad (x, y, z \in M)$.

It is easy to check that $K(M)$ is a Lie structure with $\langle x, y\rangle = xy - yx$. So is $T(M)$ with the same definition of $\langle x, y\rangle$, and other examples will occur below.

As with Lie algebras, one can easily obtain Lie structures from similar associative structures. We define an *associative structure over* M to be a $T(M)$-bimodule B together with a bilinear function $M \otimes M \to B$ (denoted by $x \otimes y \mapsto (x, y)$) satisfying the associative law:

(A) $(x, y)z = x(y, z) \quad (x, y, z \in M)$.

Then B becomes a Lie structure over M if we define $\langle x, y\rangle = (x, y) - (y, x)$. Axioms (L1) and (L3) are obviously satisfied. Axiom (L2) need only be checked when $t = z_1 z_2 \cdots z_n$ $(z_i \in M)$, and in this case axiom (A) implies $(x, y)z_1 z_2 \cdots z_n uv = xyz_1 z_2 \cdots z_n(u, v)$, from which (L2) follows easily.

For any given module M there is a universal Lie structure $L(M)$ over M which can be described as follows. $L(M)$ is generated as $T(M)$-bimodule by symbols $\langle x, y\rangle$, one for each pair of elements x, y of M, and has defining relations (L1), (L2), (L3) together with the relations that assert the bilinearity of the function $\langle,\rangle$. This $L(M)$ is characterised up to isomorphism by the universal property: if $(A, \langle\langle,\rangle\rangle)$ is any Lie structure over M then there is a unique morphism $L(M) \to A$ of $T(M)$-bimodules such that $\langle x, y\rangle \mapsto \langle\langle x, y\rangle\rangle$ for all $x, y \in M$.

There is also, of course, a universal associative structure over M, but we do not need a special notation for it since it can easily be identified. In $T(M)$, the ideal $M^2T(M) = \bigoplus_{n\geq 2}(\otimes^n M)$ is an associative structure over M with $(x, y) = xy$, and we have

Theorem 1. *$M^2T(M)$ is the universal associative structure over M.*

Proof. Let $(B, (,))$ be any associative structure over M. For $j \geq 2$ we can map $\otimes^j M$ to B by the rule $x_1 \otimes x_2 \otimes \cdots \otimes x_j \mapsto (x_1, x_2)x_3 \cdots x_j$ since the latter is an R-multilinear function. This gives a map $\theta : M^2T(M) \to B$ which is clearly a homomorphism of right $T(M)$-modules and maps xy to (x, y). To show that it is also a left $T(M)$-homomorphism it is enough to show that $y_1y_2\cdots y_n(x_1,x_2)x_3\cdots x_j = (y_1,y_2)y_3\cdots y_nx_1\cdots x_j$, $(y_r, x_s \in M)$, and this is a consequence of the associative law for B. The map θ is unique because the elements xy $(x, y \in M)$ generate $M^2T(M)$ as $T(M)$-bimodule. □

It is unfortunately not always true that $K(M)$ is the universal Lie structure over M. Indeed, it is precisely when $K(M) \cong L(M)$ that we can prove the Birkhoff–Witt theorem for Lie algebras on the module M. The proof of this theorem given in the next section is essentially that of Lazard [5], with some simplifications made possible by the axiomatic approach.

3. The Birkhoff–Witt Theorem

Suppose that we are given a multiplication $M \otimes M \to M$ $(x \otimes y \mapsto [x, y])$ which makes M a Lie algebra over R. The elements $\langle x, y\rangle = xy - yx - [x, y]$ of $T = T(M)$ $(x, y \in M)$ generate a (non-homogeneous) ideal J, and the quotient algebra $E = T/J$ is the *enveloping algebra* of the Lie algebra M. The filtration $T_0 \subset T_1 \subset T_2 \subset \cdots$ of T given by $T_n = \bigoplus_{i\leq n}(\otimes^i M)$ induces filtrations of $K = K(M)$, $S = S(M)$, J and E as follows: $K_n = K \cap T_n$, $S_n = (T_n + K)/K$, $J_n = J \cap T_n$, $E_n = (T_n + J)/J$. E is then a filtered algebra and we denote by $G = \bigoplus_{n\geq 0}(E_n/E_{n-1})$ the associated graded algebra[1]. The homogeneous components of G are

$$E_n/E_{n-1} \cong (T_n + J)/(T_{n-1} + J) \cong T_n/(T_n \cap (T_{n-1} + J)) \cong T_n/(T_{n-1} + J_n).$$

Now S is itself graded with homogeneous components

$$S_n/S_{n-1} \cong T_n/(T_{n-1} + K_n),$$

and it is clear from the definition of J that $T_{n-1} + J_n \supset T_{n-1} + K_n$. We therefore have canonical surjections $\sigma_n : S_n/S_{n-1} \to E_n/E_{n-1}$ with kernels $(T_{n-1} + J_n)/(T_{n-1} + K_n)$. These give the canonical surjection $\sigma : S \to G$ which is in fact the algebra homomorphism induced by the canonical map from M to the (commutative) algebra G. If σ is an isomorphism we say that the Lie algebra M has the *Birkhoff–Witt property*, and this is clearly equivalent to the condition

$$J_n \subset T_{n-1} + K_n \qquad (n \geq 1). \tag{1}$$

Theorem 2. *If $(K(M), xy - yx)$ is the universal Lie structure over M then every Lie algebra on M has the Birkhoff–Witt property.*

Proof. Following Lazard, we introduce modules $J_{(n)}$ $(n \geq 1)$ which consist of those elements of J that are most obviously in T_n, namely

$$J_{(n)} = \sum (T_r\langle x, y\rangle T_s \mid x, y \in M,\, r + s = n - 2).$$

Then $J_{(1)} = 0$, $J_{(2)}$ is the R-module spanned by all $\langle x, y\rangle = xy - yx - [x, y]$, and $J_{(1)} \subset J_{(2)} \subset \cdots \subset J_{(n)} \subset \cdots \subset J$. Also $J_{(n)} \subset J_n$, and $\bigcup_n J_{(n)} = J$. We write $A_n = J_{(n)}/J_{(n-1)}$ $(n \geq 2)$, and form the graded R-module $A = \bigoplus_{n\geq 2} A_n$ associated with this new filtration of J. Since $J_{(n)}M + MJ_{(n)} \subset J_{(n+1)}$, A has the structure of a T-bimodule with $A_nM + MA_n \subset A_{n+1}$. Also, $\langle x, y\rangle \in J_{(2)} = A_2$ for all $x, y \in M$.

[1] *Comment by DG:* The direct sum ranged over $n \geq 1$ in the original.

Lemma. *$(A, \langle , \rangle)$ is a Lie structure over M.*

Proof. (L1): this is trivially true since $\langle x, x\rangle = [x, x] = 0$ in M.

(L2): let $w = \langle x, y\rangle t(uv - vu) - (xy - yx)t\langle u, v\rangle$, where $x, y, u, v \in M$, and where $t \in T$ is homogeneous of degree n. Then, calculating in A, we have $w \in A_{n+4}$ since $\langle x, y\rangle \in A_2$ and $\langle u, v\rangle \in A_2$. But if we calculate in $J_{(n+4)}$ we find that

$$w = \langle x, y\rangle t\{\langle u, v\rangle + [u, v]\} - \{\langle x, y\rangle + [x, y]\}t\langle u, v\rangle = \langle x, y\rangle t[u, v] - [x, y]t\langle u, v\rangle,$$

and this lies in $J_{(n+3)}$ since $[u, v]$ and $[x, y]$ lie in M. Hence $w = 0$ in A, and (L2) follows for all t by linearity.

(L3): let $x, y, z \in M$, and put

$$u = \{\langle x, y\rangle z - z\langle x, y\rangle\} + \{\langle y, z\rangle x - x\langle y, z\rangle\} + \{\langle z, x\rangle y - y\langle z, x\rangle\}.$$

Then, calculating in A, we have $u \in A_3$. On the other hand, calculating in $J_{(3)}$, if we replace $\langle x, y\rangle$ by $xy - yx - [x, y]$, we obtain

$$u = \{[x, y]z - z[x, y]\} + \{[y, z]x - x[y, z]\} + \{[z, x]y - y[z, x]\},$$

the other terms cancelling. Now $[x, y] \in M$, so $\langle [x, y], z\rangle \in J_{(2)}$, that is,

$$[x, y]z - z[x, y] \equiv [[x, y], z] \pmod{J_{(2)}}.$$

Permuting x, y, z cyclically and adding, we therefore have $u \in J_{(2)}$ by the Jacobi law in M, and this means that $u = 0$ in A. The lemma is now proved. □

To prove the theorem we use the hypothesis that $K(M)$ is the *universal* Lie structure over M to obtain a morphism $\theta : K \to A$ of T-bimodules sending $xy - yx$ to $\langle x, y\rangle$ for all $x, y \in M$. We claim that θ is an isomorphism. For let δ_n denote the R-linear map which sends any element of T to its homogeneous part of degree n. It is clear from the definition of $J_{(n)}$ that δ_n maps $J_{(n)}$ into K_n and therefore induces a map

$$\delta_n^* : A_n = J_{(n)}/J_{(n-1)} \to K_n/K_{n-1}.$$

The maps δ_n^* combine to give a map $\delta : A \to \bigoplus(K_n/K_{n-1}) = K$ which sends $\langle x, y\rangle$ to $xy - yx$ $(x, y \in M)$, and it is easy to check that δ is a morphism of T-bimodules. Since A and K are generated as T-bimodules by all $\langle x, y\rangle$ and all $xy - yx$, respectively, we see that θ and δ are inverse isomorphisms. In particular, $\delta_n^* : J_{(n)}/J_{(n-1)} \to K_n/K_{n-1}$ is an injection and it follows that every element of $J_{(n)}$ not in $J_{(n-1)}$ has leading term of degree exactly n. Since $\bigcup_n J_{(n)} = J$, this implies that $J_{(n)} = J \cap T_n = J_n$ for all n [2]. But $J_{(n)} \subset K_n + T_{n-1}$, so we have established the condition (1) which is equivalent to the Birkhoff–Witt property. □

[2] *Comment by DG:* Here are the details of this argument: We must show that $J_{(n)} = J_n$. Since $J_{(n)} \subset J_n$ is known, it suffices to show that $J_n \subset J_{(n)}$. So let $t \in J_n$. Then, $t \in J$, so that $t \in J_{(m)}$ for some $m \geq 0$. Consider the smallest such m. Then, $t \in J_{(m)} \setminus J_{(m-1)}$, so that the residue class of t in $J_{(m)}/J_{(m-1)}$ is nonzero. Since $\delta_m^* : J_{(m)}/J_{(m-1)} \to K_m/K_{m-1}$ is an injection (as we showed above), we conclude that $\delta_m(t)$ is nonzero as well. In other words, t has a nonzero homogeneous part of degree m. Since $t \in J_n \subset T_n$, this entails $m \leq n$, so that $t \in J_{(m)} \subset J_{(n)}$. Having proved this for each $t \in J_n$, we thus obtain $J_n \subset J_{(n)}$, just as desired.

4. Lie Structures over Free Modules

Before investigating Lie structures over arbitrary modules M we need to know the situation for the special case when M is free. Our next theorem, combined with Theorem 2, gives a new proof of the Birkhoff–Witt theorem in this case.

Theorem 3. *Let M be a free R-module. Then every Lie structure $(A, \langle , \rangle)$ over M can be embedded in an associative structure $(B, (,))$ over M so that $\langle x, y\rangle = (x, y) - (y, x)$.*

Proof. Let A be a given Lie structure and put $B = A \oplus S(M)$. We shall show how to make B an associative structure with the required property. Let X be a basis for M over R and take a fixed total ordering $\leq$ of X. If $x_i \in X$ we denote by ξ_i its image in $S(M) = S$. Then S has a basis consisting of all products $\xi_1\xi_2 \cdots \xi_n$ $(n \geq 0)$, where the x_i are in X and $x_1 \leq x_2 \leq \cdots \leq x_n$.

We make B a T-bimodule as follows. The action of T on A is to be the given action. To define the action of T on S we need only define maps $M \otimes S \to B$ $(m \otimes \sigma \to m\sigma)$ and $S \otimes M \to B$ $(\sigma \otimes m \to \sigma m)$ such that $(m\sigma)n = m(\sigma n)$ for all $m, n \in M$, $\sigma \in S$. Since M is free we can define $x\sigma$ and σy arbitrarily in B for $x, y \in X$ and σ a basis element $\xi_1\xi_2 \cdots \xi_n$ of S, and we need only check that

$$(x\sigma)y = x(\sigma y) \tag{2}$$

in this case. So let $\sigma = \xi_1\xi_2 \cdots \xi_n$ $(x_1 \leq x_2 \leq \cdots \leq x_n$ in $X)$ and let ξ, η be the images of x, y in S. We define

$$x\sigma = x \circ \sigma + \xi\sigma, \qquad \sigma y = \sigma \circ y + \sigma\eta,$$

where $\xi\sigma$, $\sigma\eta$ are products in S and $x \circ \sigma$, $\sigma \circ y$ are the elements of A defined by

$$x \circ \sigma = \sum_{x_i < x} x_1x_2 \cdots x_{i-1}\langle x, x_i\rangle x_{i+1} \cdots x_n,$$
$$\sigma \circ y = \sum_{x_i > y} x_1x_2 \cdots x_{i-1}\langle x_i, y\rangle x_{i+1} \cdots x_n,$$

with the convention that $x \circ 1 = 1 \circ y = 0$. (These definitions are just an imitation of the operations in the associative structure $T(M)$ in terms of the splitting $T(M) = K(M) + S(M)$). Since $(\xi\sigma)\eta = \xi(\sigma\eta)$ in S, equation (2) is equivalent to

$$(\xi\sigma) \circ y - x(\sigma \circ y) = x \circ (\sigma\eta) - (x \circ \sigma)y. \tag{3}$$

[3] To simplify the notation, let $\tau = \xi\sigma\eta$ (product in S), and rename x, y and $x_1, x_2, \ldots, x_n$ so that $\tau = \eta_1\eta_2 \cdots \eta_k$ with[4] $y_1 \leq y_2 \leq \cdots \leq y_k$ $(k = n + 2)$ and $x = y_r$, $y = y_s$ $(r \neq s)$. Then $\sigma = (\eta_1\eta_2 \cdots \eta_k)_{\widehat{r}\widehat{s}}$, where the subscripts $\widehat{r}$, $\widehat{s}$ denote that the factors with subscripts

[3] *Comment by DG:* The following proof of (3) is rather heavy on the reader's concentration. An alternative proof, found by GPT-5.5, is included in the Appendix below.

[4] *Comment by DG:* The η_i are understood to be images of the y_i in $S(M) = S$.

r, s are to be omitted. Also $\xi\sigma = (\eta_1\eta_2\cdots\eta_k)_{\widehat{s}}$ and $\sigma\eta = (\eta_1\eta_2\cdots\eta_k)_{\widehat{r}}$. Writing[5] $\epsilon_{pq} = 0$ if $p \le q$ and $\epsilon_{pq} = 1$ if $p > q$, the left hand side of (3) becomes

$$\begin{aligned}(\xi\sigma)\circ y - x(\sigma\circ y) = &\sum_j \epsilon_{js}(y_1\cdots y_{j-1}\langle y_j, y_s\rangle y_{j+1}\cdots y_k)_{\widehat{s}}\\ &- y_r\sum_{j\neq r}\epsilon_{js}(y_1\cdots y_{j-1}\langle y_j, y_s\rangle y_{j+1}\cdots y_k)_{\widehat{r}\widehat{s}}.\end{aligned}$$

The terms $j \neq r$ in the first sum appear in the second sum with y_r moved to the left hand end. This move can be accomplished by adding terms containing commutators $y_iy_r - y_ry_i$ and possibly $\langle y_j, y_s\rangle y_r - y_r\langle y_j, y_s\rangle$. We therefore have $(\xi\sigma)\circ y - x(\sigma\circ y) = U + V + W$, where

$$\begin{aligned}U &= \epsilon_{rs}(y_1\cdots y_{r-1}\langle y_r, y_s\rangle y_{r+1}\cdots y_k)_{\widehat{s}},\\ V &= -\sum_{j\neq r}\sum_{\substack{i\neq s\\ i\neq j}}\epsilon_{js}\epsilon_{ri}(y_1\cdots y_{i-1}(y_ry_i - y_iy_r)y_{i+1}\cdots y_{j-1}\langle y_j, y_s\rangle y_{j+1}\cdots y_k)_{\widehat{r}\widehat{s}},\\ W &= \sum_j \epsilon_{js}\epsilon_{rj}(y_1\cdots y_{j-1}(\langle y_j, y_s\rangle y_r - y_r\langle y_j, y_s\rangle)y_{j+1}\cdots y_k)_{\widehat{r}\widehat{s}}.\end{aligned}$$

(The notation in V is not meant to imply that $i < j$). Similarly, the right hand side of (3) is $U' + V' + W'$, where

$$\begin{aligned}U' &= \epsilon_{rs}(y_1\cdots y_{s-1}\langle y_r, y_s\rangle y_{s+1}\cdots y_k)_{\widehat{r}},\\ V' &= -\sum_{i\neq s}\sum_{\substack{j\neq r\\ j\neq i}}\epsilon_{ri}\epsilon_{js}(y_1\cdots y_{i-1}\langle y_r, y_i\rangle y_{i+1}\cdots y_{j-1}(y_jy_s - y_sy_j)y_{j+1}\cdots y_k)_{\widehat{r}\widehat{s}},\\ W' &= \sum_j \epsilon_{rj}\epsilon_{js}(y_1\cdots y_{j-1}(y_s\langle y_r, y_j\rangle - \langle y_r, y_j\rangle y_s)y_{j+1}\cdots y_k)_{\widehat{r}\widehat{s}}.\end{aligned}$$

Now $V = V'$ by axiom (L2) in A. Also, by axiom (L3),

$$\begin{aligned}W - W' &= \sum_{s<j<r}(y_1\cdots y_{j-1}(y_j\langle y_s, y_r\rangle - \langle y_s, y_r\rangle y_j)y_{j+1}\cdots y_k)_{\widehat{r}\widehat{s}}\\ &= \epsilon_{rs}\{(y_1\cdots y_{r-1}\langle y_s, y_r\rangle y_{r+1}\cdots y_k)_{\widehat{s}}\\ &\qquad - (y_1\cdots y_{s-1}\langle y_s, y_r\rangle y_{s+1}\cdots y_k)_{\widehat{r}}\}\\ &= U' - U\end{aligned}$$

since (L1) implies that $\langle y_s, y_r\rangle = -\langle y_r, y_s\rangle$. Thus equation (3) is satisfied and B is a T-bimodule.

Now for $x, y \in X$ we have

$$x\eta = \xi y = \begin{cases}\xi\eta & \text{if } x \le y\\ \xi\eta + \langle x, y\rangle & \text{if } x \ge y.\end{cases} \tag{4}$$

[5] *Comment by DG:* The "p" and "q" in this notation were called "r" and "s" in the original. I renamed them to disambiguate them from the r and s already defined.

We may therefore define $(x, y) = x\eta = \xi y$ for $x, y \in X$, and extend linearly to the whole of M. Then B is an associative structure over M since the equation $(x, y)z = x(y, z)$ is R-multilinear and is a special case of (2) when $x, y, z \in X$. Finally, $(x, y) - (y, x) = x\eta - y\xi = \langle x, y\rangle$ by (4) when $x, y \in X$, and the equality holds for $x, y \in M$ by linearity. □

Corollary 1. *If M is a free R-module then $K(M)$ is the universal Lie structure over M.*

Proof. Let $(A, \langle , \rangle)$ be any Lie structure over M and embed A in the associative structure B as above. By Theorem 1 there is a map $\theta : M^2T(M) \to B$ of $T(M)$-bimodules sending xy to (x, y) for $x, y \in M$. Since $(x, y) - (y, x) = \langle x, y\rangle \in A$, θ induces a map of $T(M)$-bimodules $K(M) \to A$ sending $xy - yx$ to $\langle x, y\rangle$. □

Using Theorem 2 we now obtain

Corollary 2. *Every Lie algebra over R whose underlying module is free has the Birkhoff–Witt property.*

5. Baer Invariants of Tensor Algebras

We now introduce two invariants of a module M analogous to the invariants $[F, F]/[F, R]$ and $(R \cap [F, F])/[F, R]$ for a group presented as a quotient F/R of a free group. In Fröhlich's notation [4] they are $D_0V(T(M))$ and $D_1U(T(M))$, where U and V are the functors on algebras associated with the variety of commutative algebras: $U(T(M)) = S(M)$, $V(T(M)) = K(M)$. We recall briefly their definition and main properties.

We start with a short exact sequence of R-modules

$$0 \to Q \to P \to M \to 0 \tag{5}$$

with P projective, and we make identifications so that $Q \subset P \subset T(P)$. Then we have a commutative diagram with exact rows and columns

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & K(P) \cap \overline{Q} & \longrightarrow & \overline{Q} & \longrightarrow & \overline{Q}^* & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & K(P) & \longrightarrow & T(P) & \longrightarrow & S(P) & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & K(M) & \longrightarrow & T(M) & \longrightarrow & S(M) & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array}$$

where $\overline{Q}$ is the ideal $T(P)QT(P)$ of $T(P)$ generated by Q, and $\overline{Q}^*$ is its image in $S(P)$ [6]. Let W be the set of words

$$w_i(\mathbf{x}) = w_i(x_1, x_2, \ldots, x_n) = x_1 x_2 \cdots x_{i-1}[x_i, x_{i+1}]x_{i+2} \cdots x_n$$

($n \geq 2$, $1 \leq i \leq n-1$), where $[a, b]$ denotes, as always from now on, the additive commutator $ab - ba$. If the x_j take values running through P then W generates $K(P)$ as additive group. Let W' be the set of derived words $w'_i(\mathbf{x}, \mathbf{y}) = w_i(\mathbf{x} + \mathbf{y}) - w_i(\mathbf{x})$. If the x_j run through P and the y_j run through Q, W' generates (as additive group) an R-module Z. Clearly the image of any such element $w'_i(\mathbf{p}, \mathbf{q})$ in $T(M)$ or in $S(P)$ is 0, so $Z \subset K(P) \cap \overline{Q}$, and it is not difficult to check that in fact Z is the ideal of $T(P)$ generated by all $[t, q]$ with $t \in T(P)$

[6] *Comment by DG:* The exactness of the rows is obvious. The exactness of the second and third columns follows from known results (Theorem 32 and Theorem 62 in Darij Grinberg, *A few classical results on tensor, symmetric and exterior powers*, 17 April 2026, to name just one reference). The exactness of the first column follows from the exactness of the second.

and $q \in \overline{Q}$ [7]. We write

$$B(M) = (K(P) \cap \overline{Q})/Z \qquad \text{and} \qquad C(M) = K(P)/Z.$$

(These are respectively $D_1UT(M)$ and $D_0VT(M)$).

Now $K(P)$ and $\overline{Q}$ and Z are $T(P)$-bimodules and satisfy[8] $K(P)\overline{Q}+\overline{Q}K(P) \subset Z$ (because e.g. $p_1p_2 \cdots [p_i, p_{i+1}] \cdots q \cdots p_n$ is of the form $w_i'(\mathbf{p}, \mathbf{q})$). Hence $B(M)$ and $C(M)$ are $T(M)$-bimodules, and we have an exact sequence of $T(M)$-bimodules:

$$0 \to B(M) \to C(M) \to K(M) \to 0. \tag{6}$$

As the notation suggests, $B(M)$ and $C(M)$ depend only on M and not on its presentation (5). To see this directly, let $0 \to Q' \to P' \to M \to 0$ be another presentation of M with P'

[7] *Comment by DG:* In the original: "$Z = [T(P), \overline{Q}]$, the ideal of $T(P)$ generated by all $[t, q]$ ($t \in T(P)$, $q \in \overline{Q}$)". But the notation $[T(P), \overline{Q}]$ for this is rather nonstandard; usually it would mean just the R-linear span of all commutators $[t, q]$ with $t \in T(P)$ and $q \in \overline{Q}$, and this span does not equal Z.

Here is a brief outline of the proof: It is easy to see from the definition that

$$Z = \operatorname{span}\{w_i(x_1, x_2, \ldots, x_n) \mid \text{each } x_j \text{ belongs to } P, \text{ and at least one } x_j \text{ belongs to } Q\}.$$

(Indeed, this is perhaps the better definition of Z.) This shows that Z is an ideal of $T(P)$. Now, let $\widetilde{Z}$ be the ideal of $T(P)$ generated by all $[t, q]$ with $t \in T(P)$ and $q \in \overline{Q}$. We must prove that $Z = \widetilde{Z}$. We shall achieve this by proving both $\widetilde{Z} \subseteq Z$ and $Z \subseteq \widetilde{Z}$:

1. To prove $\widetilde{Z} \subseteq Z$, we need to show that every $t \in T(P)$ and $q \in \overline{Q}$ satisfy $[t, q] \in Z$ (since Z is an ideal of $T(P)$). By linearity, we can assume that $t = x_1x_2 \cdots x_n$ with $x_1, x_2, \ldots, x_n \in P$, and that $q = aq'b$ with $a = y_1y_2 \cdots y_k \in T(P)$ and $b = y_{k+1}y_{k+2} \cdots y_\ell \in T(P)$ and $q' \in Q$. But then we have

   $$[t, q] = [t, aq'b] = [t, a]q'b + a[t, q']b + aq'[t, b]$$

   and

   $$[t, q'] = [x_1x_2 \cdots x_n, q'] = \sum_i x_1x_2 \cdots x_{i-1}[x_i, q']x_{i+1}x_{i+2} \cdots x_n.$$

   Substituting the latter equality into the former, we obtain

   $$[t, q] = [t, a]q'b + a\left(\sum_i x_1x_2 \cdots x_{i-1}[x_i, q']x_{i+1}x_{i+2} \cdots x_n\right) b + aq'[t, b].$$

   All addends on the right hand side (once expanded) are of the form $w_i(x_1, x_2, \ldots, x_n)$ where each x_j belongs to P and at least one x_j belongs to Q. Thus, they belong to Z. Hence, $[t, q] \in Z$. So we have shown that $\widetilde{Z} \subseteq Z$.

2. In order to prove the reverse inclusion $Z \subseteq \widetilde{Z}$, we must show that each $w_i(x_1, x_2, \ldots, x_n)$ with all $x_j \in P$ and at least one $x_j \in Q$ belongs to $\widetilde{Z}$. This is clear if the one x_j that belongs to Q is x_i or x_{i+1}, because in this case the $[x_i, x_{i+1}]$ bracket in the middle of $w_i(x_1, x_2, \ldots, x_n)$ itself has the form $[t, q]$ with $t \in T(P)$ and $q \in \overline{Q}$ (actually, $q = x_j \in Q$). In all other cases, we first move this one x_j to a position either just left or just right of the $[x_i, x_{i+1}]$ bracket by successively commuting it past the other x_k's (the commutators will lie in $\widetilde{Z}$), and then use $x_j[x_i, x_{i+1}] = [x_jx_i, x_{i+1}] - [x_j, x_{i+1}]x_i$ (if x_j ended up left of the bracket) or $[x_i, x_{i+1}]x_j = [x_i, x_{i+1}x_j] - x_{i+1}[x_i, x_j]$ (if on the right), obtaining a right hand side belonging to $\widetilde{Z}$ in both cases (since x_j, x_jx_i and $x_{i+1}x_j$ all belong to $\overline{Q}$).

[8] *Comment by DG:* Added "and Z" as well as a few filler words (to avoid overfull hbox).

projective. Then there are R-linear maps σ, σ_* making the following diagram commute:

$$\begin{array}{ccccccccc}
0 & \longrightarrow & Q & \longrightarrow & P & \longrightarrow & M & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle \sigma_*} & & \downarrow{\scriptstyle \sigma} & & \downarrow{\scriptstyle 1_M} & & \\
0 & \longrightarrow & Q' & \longrightarrow & P' & \longrightarrow & M & \longrightarrow & 0.
\end{array}$$

The map $T(\sigma) : T(P) \to T(P')$ is an algebra homomorphism and sends $\overline{Q}$ into $\overline{Q'}$. Also, since $K(P)$ and Z are defined by algebra words, $T(\sigma)$ sends $K(P)$ into $K(P')$ and Z into Z'. Hence σ induces maps $B(M) \to B'(M)$ and $C(M) \to C'(M)$ (in the obvious notation). Similarly, there is a map $\tau : P' \to P$ inducing maps in the opposite direction. It is enough, therefore, to show that if $\rho : P \to P$ induces the identity map on M then it induces the identity map on $B(M)$ and on $C(M)$. But this is clear since if $u \in K(P)$ then u is a sum of elements $w_i(\mathbf{p})$ ($p_j \in P$), and so $u\rho - u = \sum(w_i(\mathbf{p}\rho) - w_i(\mathbf{p})) \in Z$ (because $p_j\rho - p_j \in Q$). A similar argument with 1_M replaced by an arbitrary map of R-modules shows that B and C are functors from R-modules to R-modules[9].

We now show that $C(M) = K(P)/Z$ is a Lie structure over M in a natural way. We have already shown that $C(M)$ is a $T(M)$-bimodule. We also know that $K(P)$ is a Lie structure over P with respect to the operation $[x, y] = xy - yx$ ($x, y \in P$). Suppose that $x \equiv x'$ (mod Q) and $y \equiv y'$ (mod Q). Then $[x, y] \equiv [x', y']$ (mod Z), so for $\xi, \eta \in M$ we may define $[\![\xi, \eta]\!] \in C(M) = K(P)/Z$ to be the image in $C(M)$ of $[x, y]$, where $x, y \in P$ have images ξ, η in M. The axioms for a Lie structure hold in $C(M)$ over M because they hold in $K(P)$ over P.

Theorem 4. *For any R-module M, $(C(M), [\![,]\!])$ is the universal Lie structure over M.*

Proof. Let $(A, \langle, \rangle)$ be any Lie structure over M. In constructing $C(M)$ we may choose P to be a free R-module, and in this case we know (Theorem 3, Corollary 1) that $(K(P), [,])$ is the universal Lie structure over P. Now A can be viewed as a Lie structure over P *via* the map

[9] *Comment by DG:* Let me expand upon this somewhat. Let me show that B is a functor (the proof for C is analogous). Each R-module M has a canonical presentation $0 \to Q_M \to P_M \to M \to 0$, where P_M is the free R-module with M as a basis, where $P_M \to M$ is the canonical R-module epimorphism that sends each basis element corresponding to some $m \in M$ to the respective m itself, and where Q_M is the kernel of this epimorphism. We can define $B(M)$ canonically (not just up to isomorphism) using this presentation. To show that this gives a functor, we define $B(f) : B(M) \to B(N)$ for any R-linear map $f : M \to N$ as follows: Lift f to an R-linear map $P_f : P_M \to P_N$ (this lift is not canonical, but exists because P_M is free and thus projective). Then, P_f sends Q_M to Q_N, and thus we obtain a commutative diagram

$$\begin{array}{ccccccccc}
0 & \longrightarrow & Q_M & \longrightarrow & P_M & \longrightarrow & M & \longrightarrow & 0 \\
 & & {\scriptstyle Q_f}\downarrow & & {\scriptstyle P_f}\downarrow & & {\scriptstyle f}\downarrow & & \\
0 & \longrightarrow & Q_N & \longrightarrow & P_N & \longrightarrow & N & \longrightarrow & 0.
\end{array}$$

This leads to an algebra homomorphism $T(P_f) : T(P_M) \to T(P_N)$ that sends $K(P_M)$ to $K(P_N)$, sends Z_M to Z_N, and sends $\overline{Q}_M$ to $\overline{Q}_N$, and thus gives rise to an R-linear map $B(f) : B(M) \to B(N)$. It remains to show that this $B(f)$ depends only on f and not on its lift P_f (functoriality will then follow immediately, since $T(P_f)$ is functorial in P_f). But this is clear since, if P_{f1} and P_{f2} are two candidates for the lift P_f, then each element $u \in K(P_M)$ is a sum of elements $w_i(\mathbf{p})$ with $p_j \in P_M$, and so $u\,T(P_{f1}) - u\,T(P_{f2}) = \sum(w_i(\mathbf{p}\,P_{f1}) - w_i(\mathbf{p}\,P_{f2})) \in Z$ (because $p_j\,P_{f1} - p_j\,P_{f2} \in Q_N$).

$\theta : P \to M$ of the presentation, so there is a unique map $\alpha : K(P) \to A$ of $T(P)$-bimodules sending $[x, y]$ to $\langle x\theta, y\theta\rangle$ for all $x, y \in P$. If $x_1, x_2, \ldots, x_n \in P$ then

$$\begin{aligned} w_i(x_1, x_2, \ldots, x_n)\alpha &= (x_1 \cdots x_{i-1}[x_i, x_{i+1}]x_{i+2} \cdots x_n)\alpha \\ &= (x_1\theta) \cdots (x_{i-1}\theta)\langle x_i\theta, x_{i+1}\theta\rangle(x_{i+2}\theta) \cdots (x_n\theta). \end{aligned}$$

Hence, for $q_1, q_2, \ldots, q_n \in Q$, $(w_i(\mathbf{x} + \mathbf{q}) - w_i(\mathbf{x}))\alpha = 0$, i.e. $Z \subset \operatorname{Ker}\alpha$. Thus α induces a map $\beta : C(M) = K(P)/Z \to A$ [10] sending $[\![\xi, \eta]\!]$ to $\langle \xi, \eta\rangle$ for all $\xi, \eta \in M$. It is easy to see that β is a morphism of $T(M)$-bimodules, and it is unique since the $[x, y]$ generate $K(P)$ as $T(P)$-bimodule and therefore the $[\![\xi, \eta]\!]$ generate $C(M)$ as $T(M)$-bimodule. □

The canonical maps $C(M) \to L(M)$ given by this theorem form a natural equivalence of functors $C \simeq L$. We may therefore write $L(M)$ for $C(M)$ from now on, and we have the exact sequence

$$0 \to B(M) \to L(M) \to K(M) \to 0 \tag{6'}$$

for all R-modules M. Clearly this gives an exact sequence of functors $0 \to B \to L \to K \to 0$, and combining it with $0 \to K \to T \to S \to 0$ we obtain the exact sequence of functors

$$0 \to B \to L \to T \to S \to 0. \tag{7}$$

Theorem 5. *For any R-module M the following are equivalent:*

(i) $B(M) = 0$;

(ii) $(K(M), [,])$ *is the universal Lie structure over* M;

(iii) every Lie structure over M is embeddable in an associative structure over M.

Proof. The equivalence of (i) and (ii) follows from the exact sequence (6′) (which is a consequence of Theorem 4).

(ii) ⇒ (iii). Let $(A, \langle,\rangle)$ be any Lie structure over M and let A_0 be the $T(M)$-bimodule generated in A by the elements $\langle x, y\rangle$ $(x, y \in M)$. If (ii) holds then there is a unique morphism $\alpha : K(M) \to A$ of $T(M)$-bimodules sending $[x, y]$ to $\langle x, y\rangle$. The kernel D of α is a $T(M)$-bimodule, i.e. an ideal of $T(M)$. The algebra $A^* = T(M)/D$ is an associative structure over M and the Lie structure $A_0 \cong K(M)/D$ is embedded in it. To extend this to an embedding of A itself is a trivial matter. Any $T(M)$-bimodule containing A^* is also an associative structure over M, so we need only form the fibre coproduct of A and A^* with respect to the embeddings $A_0 \to A$ and $A_0 \to A^*$. It is clear that $\langle x, y\rangle$ goes to $xy - yx$ in the resulting embedding of A.

(iii) ⇒ (ii). If (iii) holds then the universal Lie structure $L = (L(M), \langle,\rangle)$ is embeddable in an associative structure $(L^*, (,))$ so that $(x, y) - (y, x) = \langle x, y\rangle$. By Theorem 1, there is a morphism of $T(M)$-bimodules $\alpha : M^2T(M) \to L^*$ sending xy to (x, y), and this induces a morphism $\beta : K(M) \to L$ sending $xy - yx$ to $\langle x, y\rangle$. Clearly β is inverse to the canonical map $L(M) \to K(M)$, so $K(M) \cong L(M)$. □

[10] *Comment by DG:* The original says "$\beta : C(M) = K(M)/Z \to A$". Claude corrected this autonomously.

6. Modules with $B(M) = 0$

Our main result is an immediate consequence of Theorems 2 and 5:

Theorem 6. *If $B(M) = 0$ for the R-module M then the Birkhoff–Witt theorem holds for all Lie algebras over R with underlying module M.*

To show that this theorem contains the known results quoted in the introduction we now look for conditions on the R-module M which ensure that $B(M) = 0$.

Theorem 7. *Let R be a fixed commutative ring and let M be any R-module.*

(i) If M is R-projective then $B(M) = 0$.

(ii) If M is uniquely divisible as Abelian group (i.e. M is a rational vector space) then $B(M) = 0$.

(iii) If M is a direct sum of cyclic (i.e. one-generator) modules then $B(M) = 0$.

Proof. Let $0 \to Q \to P \to M \to 0$ be a presentation of M with P projective. Then, in the notation of Section 5, $B(M) = (K(P) \cap \overline{Q})/Z$. Item (i) is clear since if M is projective we may take $P = M$ and $Q = 0$. To prove (ii) and (iii) we first observe that $K(P)$, $\overline{Q}$ and Z are homogeneous ideals of $T(P)$, so it is enough to take $u \in K(P) \cap \overline{Q}$ homogeneous of degree n ($n \geq 2$) and show that $u \in Z$. Now the symmetric group $\mathscr{S}_n$ acts on the homogeneous part $\otimes^n P$ of $T(P)$, and if $\pi \in \mathscr{S}_n$, $u \in \otimes^n P$, then $u - u\pi \in K(P)$. Moreover, if $u \in \overline{Q}$, then $u - u\pi$ is a sum of elements of type $p_1 p_2 \cdots p_{i-1}[p_i, q]p_{i+2} \cdots p_n$ or $p_1 \cdots p_{i-1}[p_i, p_{i+1}]p_{i+2} \cdots q \cdots p_n$, where $p_j \in P$ and $q \in Q$. All such elements are in Z, by definition, so $u - u\pi \in Z$ whenever $u \in \overline{Q} \cap \otimes^n P$. On the other hand, if $u \in K(P)$ and is homogeneous of degree n then u is a sum of elements of the form $v - v\tau$, where $\tau \in \mathscr{S}_n$ is a transposition. Hence $\sum_{\pi \in \mathscr{S}_n} u\pi = \sum_{\pi \in \mathscr{S}_n}(v\pi - v\tau\pi) = 0$ in this case. Thus, for any $u \in \overline{Q} \cap K(P) \cap \otimes^n P$, we have $n!u = \sum_{\pi \in \mathscr{S}_n}(u - u\pi) \in Z$. This shows that $B(M)$ is always a torsion group. It is graded by degree: $B(M) = \bigoplus B^n(M)$, and $n!B^n(M) = 0$.

Suppose now that M is uniquely divisible. Then for each integer $k > 0$ we have an isomorphism $k : M \to M$ ($x \mapsto kx$). Since B is a functor this induces an isomorphism $B(k) : B(M) \to B(M)$ which in dimension n is multiplication by k^n. Taking $k = n!$ we see that $B^n(M) = (n!)^n B^n(M) = 0$ [11], which proves (ii).

To prove (iii) we suppose that $M = \bigoplus_{x \in X} M_x$, where $M_x = R/I_x$ is cyclic, and we take P to be the free R-module on X with the obvious map $P \to M$. Then Q is spanned by certain elements of the form λx, where $\lambda \in R$ and $x \in X$. We take a fixed total ordering $\leq$ of X and denote by S^* the R-submodule of $T(P)$ spanned by all products $x_1 x_2 \cdots x_n$ with $x_i \in X$ and $x_1 \leq x_2 \leq \cdots \leq x_n$ ($n \geq 0$). Then there is an R-linear map $\theta : T(P) \to S^*$ which sends any product of x's to[12] the product of the same x's in correct order. The kernel of θ is exactly $K(P)$. Suppose now that $u \in \overline{Q} \cap \otimes^n P$. Because of the special form of our presentation we have $u = u_1 + u_2 + \cdots + u_k$ where each u_i is still in $\overline{Q}$ and is of the form $\lambda x_1 x_2 \cdots x_n$ with $\lambda \in R$ and $x_1, x_2, \ldots, x_n \in X$. Then $u_i\theta = u_i\pi_i$ for some $\pi_i \in \mathscr{S}_n$, so $u_i - u_i\theta = u_i - u_i\pi_i \in Z$,

[11] *Comment by DG:* Claude fixed a typo here ("$B^n(m)$").
[12] *Comment by DG:* In the original "to be".

as we have already shown. Hence $u - u\theta \in Z$. If now $u \in \overline{Q} \cap K(P) \cap \otimes^n P$ then $u\theta = 0$, and we have $u \in Z$ as required. □

Corollary. *If R is the direct sum of a finite number of fields or is an algebra over the rationals then $B(M) = 0$ for all R-modules M. If R is a principal ideal domain then $B(M) = 0$ for all finitely generated R-modules M.*

We can extend this last result by general arguments as follows.

Theorem 8. *If $\{M_\alpha\}$ is a directed system of R-modules, and $M = \varinjlim M_\alpha$, then $B(M) = \varinjlim B(M_\alpha)$.*

Proof. The exact sequence of functors (7) gives rise to a directed system of exact sequences

$$0 \to B(M_\alpha) \to L(M_\alpha) \to T(M_\alpha) \to S(M_\alpha) \to 0.$$

Since $\varinjlim$ is an exact functor for R-modules, we obtain a commutative diagram

$$\begin{array}{ccccccccc} 0 \longrightarrow & \varinjlim B(M_\alpha) & \longrightarrow & \varinjlim L(M_\alpha) & \longrightarrow & \varinjlim T(M_\alpha) & \longrightarrow & \varinjlim S(M_\alpha) & \longrightarrow 0 \\ & \downarrow & & \downarrow{\scriptstyle\lambda} & & \downarrow{\scriptstyle\tau} & & \downarrow{\scriptstyle\sigma} & \\ 0 \longrightarrow & B(M) & \longrightarrow & L(M) & \longrightarrow & T(M) & \longrightarrow & S(M) & \longrightarrow 0 \end{array}$$

with exact rows, the limits in the upper row being taken in the category of R-modules. It is enough, therefore, to show that λ, τ, σ are isomorphisms. We give the proof for λ; the other cases are proved by similar "general nonsense" and are in any case well known.

$L_\alpha = L(M_\alpha)$ is the universal Lie structure over M_α. Its structure is given by canonical maps $M_\alpha \otimes M_\alpha \to L_\alpha$, $L_\alpha \otimes M_\alpha \to L_\alpha$ and $M_\alpha \otimes L_\alpha \to L_\alpha$ satisfying axioms (L1), (L2), (L3). If $\theta : M_\alpha \to M_\beta$ is R-linear then $\theta^* = L(\theta) : L_\alpha \to L_\beta$ is obtained by viewing L_β as a Lie structure over M_α *via* the map θ. It is therefore not only R-linear but is compatible with the structure maps, that is, for $x, y \in M_\alpha$ and $a \in L_\alpha$ we have[13] $\langle x, y\rangle\theta^* = \langle x\theta, y\theta\rangle$, $(ax)\theta^* = (a\theta^*)(x\theta)$ and $(xa)\theta^* = (x\theta)(a\theta^*)$. Hence, writing $\Lambda = \varinjlim L_\alpha$, the structure maps of the various L_α induce maps

$$M \otimes M = \varinjlim(M_\alpha \otimes M_\alpha) \to \Lambda, \qquad \Lambda \otimes M = \varinjlim(L_\alpha \otimes M_\alpha) \to \Lambda$$

and $M \otimes \Lambda \to \Lambda$. The axioms (L1), (L2), (L3) carry over to the limits since each axiom involves only a finite number of symbols and therefore each instance of it is implied by the corresponding axiom for some pair M_α, L_α. Thus Λ is in a canonical way a Lie structure over M. If now A is any Lie structure over M then A can be viewed as a Lie structure over M_α. Hence there is a unique morphism of Lie structures $L_\alpha \to A$ for each α. These induce a unique morphism $\Lambda \to A$ of Lie structures over M which, in the particular case $A = L(M)$, is the map λ. The standard argument for universal objects now shows that λ is an isomorphism. □

Corollary. *If R is a principal ideal domain then $B(M) = 0$ for all R-modules M.*

[13] *Comment by DG:* Missing subscripts inserted in "$x, y \in M$ and $a \in L$".

Finally, we consider change-of-ring arguments. If R' is a commutative R-algebra and M is an R-module then $M' = M \otimes_R R'$ is an R'-module and we may form its Baer invariant as such. We write $B_{R'}(M')$ to indicate that we are calculating with R'-modules.

Theorem 9. *If the R-algebra R' is flat over R then $B_{R'}(M \otimes_R R') = B_R(M) \otimes_R R'$.*

Proof. The argument is similar to the one given for direct limits. Since R' is flat over R we have, for any R-module M, an exact sequence of R'-modules

$$0 \to B_R(M) \otimes_R R' \to L_R(M) \otimes_R R' \to T_R(M) \otimes_R R' \to S_R(M) \otimes_R R' \to 0.$$

We write $M' = M \otimes_R R'$, $L = L_R(M)$, $L' = L_R(M) \otimes_R R'$. The structure maps for L, namely the maps[14] $M \otimes_R M \to L$, $L \otimes_R M \to L$, $M \otimes_R L \to L$, induce R'-linear maps $M' \otimes_{R'} M' \to L'$ etc. which clearly make L' a Lie structure over M', and it is easy to check that L' is then the universal Lie structure $L_{R'}(M')$. Similarly, we may identify $T_R(M) \otimes_R R'$ with $T_{R'}(M')$ and $S_R(M) \otimes_R R'$ with $S_{R'}(M')$, and the theorem follows. $\square$

In particular, the local ring $R_\mathfrak{p}$ at a prime ideal $\mathfrak{p}$ of R is flat over R (see, for example, Nagata [6], (6.18) on[15] p. 19). Writing $M_\mathfrak{p}$ for $M \otimes_R R_\mathfrak{p}$ we therefore have the following.

Corollary 1. *For any R-module M and any prime ideal $\mathfrak{p}$ of R, $B_{R_\mathfrak{p}}(M_\mathfrak{p}) = (B_R(M))_\mathfrak{p}$. Hence $B_R(M) = 0$ if and only if $B_{R_\mathfrak{p}}(M_\mathfrak{p}) = 0$ for all prime ideals (or all maximal ideals) $\mathfrak{p}$.*

Since the local rings of a Dedekind domain are principal ideal domains, the corollary to Theorem 8 now gives

Corollary 2. *If R is a Dedekind domain then $B(M) = 0$ for all R-modules M.*

Appendix 1 (by DG). A cleaner proof of equation (3)

This appendix gives a less index-heavy proof of equation (3) from Section 4. This proof was suggested by GPT-5.5, and rewritten by myself in a more elementary language. It was originally inspired by the theory of rewrite systems [8, 11].

We keep the notation from the proof of Theorem 3. Thus X is a totally ordered basis of the free R-module M, and we have defined elements $x \circ \sigma$ and $\sigma \circ y$ of A for all $x, y \in X$ and all basis monomials $\sigma = \xi_1 \xi_2 \cdots \xi_n$ of $S(M)$.

A *word* shall mean a finite list of elements of X. We shall write a word $(w_1, w_2, \ldots, w_n)$ as $w_1 w_2 \cdots w_n$, thus visually identifying it with the basis element of $T(M)$ it corresponds to.

The *inversions* of a word $w_1 w_2 \cdots w_n$ are the pairs (i, j) of integers satisfying $1 \leq i < j \leq n$ and $w_i > w_j$. In other words, they are the pairs of positions in the word such that the earlier position contains a (strictly) larger letter than the later one. We let $\operatorname{inv} w$ denote the number of inversions of a word $w = w_1 w_2 \cdots w_n$. This is a nonnegative integer.

[14] *Comment by DG:* Added "namely the maps" to avoid a false impression that had me confused too.
[15] *Comment by DG:* Added "(6.18) on" to avoid relying on page numbers.

If $w = w_1 w_2 \cdots w_n$ is a word and $1 \leq i < n$ an integer satisfying $w_i > w_{i+1}$, then we can swap the i-th and $(i+1)$-st letters of w to obtain a new word

$$w^{\frown i} := (w_1 w_2 \cdots w_{i-1})\, w_{i+1} w_i\, (w_{i+2} w_{i+3} \cdots w_n)\,,$$

which has one fewer inversion than w; that is,

$$\operatorname{inv}\left(w^{\frown i}\right) = \operatorname{inv} w - 1. \tag{8}$$

We shall say that $w^{\frown i}$ is obtained from w by a *single sorting step* in this case. Often, several different single sorting steps can be applied to a given word w; for example, from the word $w = 4231$ (over the totally ordered set $X = \mathbb{Z}$), we can obtain either $w^{\frown 1} = 2431$ or $w^{\frown 3} = 4213$ by a single sorting step.

Now, for any word $w = w_1 w_2 \cdots w_n$, we want to define recursively an element $J(w) \in A$ as follows:

1. If $\operatorname{inv} w = 0$ (that is, if w has no inversions), then we set $J(w) = 0$ in A.
2. If $\operatorname{inv} w > 0$, then there exists some integer $1 \leq i < n$ satisfying $w_i > w_{i+1}$ (since otherwise, we would have $w_1 \leq w_2 \leq \cdots \leq w_n$ and thus $\operatorname{inv} w = 0$). We choose such an i and set

$$J(w) = w_1 w_2 \cdots w_{i-1} \langle w_i, w_{i+1} \rangle\, w_{i+2} w_{i+3} \cdots w_n + J\left(w^{\frown i}\right). \tag{9}$$

   where (as we recall) $w^{\frown i} = (w_1 w_2 \cdots w_{i-1})\, w_{i+1} w_i\, (w_{i+2} w_{i+3} \cdots w_n)$ is the word obtained from w by a single sorting step swapping the i-th and $(i+1)$-st letters.

It is not obvious that this is well-defined, because in the $\operatorname{inv} w > 0$ case, there might be several different integers $1 \leq i < n$ satisfying $w_i > w_{i+1}$, and then the formula (9) might perhaps give different results depending on which of these i is chosen. We shall soon see that all choices of i lead to the same $J(w)$ – so that $J(w)$ really is well-defined – but this must be proved. What is clear is that there is always at least one candidate for $J(w)$, because the recursion is well-founded (thanks to (8)).

We now want to show that all choices of i do lead to the same $J(w)$ in (9). For now, let us modify the definition so that the equation (9) is used to define $J(w)$ only when i is the **smallest** integer $1 \leq i < n$ satisfying $w_i > w_{i+1}$. This way, the uniqueness of $J(w)$ is obvious (since the smallest i is always unique). But we must now prove that (9) holds for **each** integer $1 \leq i < n$ satisfying $w_i > w_{i+1}$, not just for the smallest.

Proof of (9). We shall prove this by strong induction on $\operatorname{inv} w$. The *base case* ($\operatorname{inv} w = 0$) is obvious, since (9) does not apply to this case at all. For the *induction step*, we pick some positive integer k, and we assume that (9) is proved for all words w with $\operatorname{inv} w < k$ (and all i that satisfy $w_i > w_{i+1}$, not just for the smallest such i). Now we shall prove the same for all words w with $\operatorname{inv} w = k$.

Fix such a word $w = w_1 w_2 \cdots w_n$. Let j be the **smallest** integer $1 \leq i < n$ satisfying $w_i > w_{i+1}$. Then, by our (modified) definition of $J(w)$, we have

$$J(w) = w_1 w_2 \cdots w_{j-1} \langle w_j, w_{j+1} \rangle\, w_{j+2} w_{j+3} \cdots w_n + J\left(w^{\frown j}\right). \tag{10}$$

Now, pick **any** integer $1 \leq i < n$ satisfying $w_i > w_{i+1}$. We must prove (9). If $i = j$, then this follows from (10); thus, we assume that $i \neq j$ from now on. Since j was the smallest i satisfying

$w_i > w_{i+1}$, we have $i \geq j$, and thus $i > j$ (since $i \neq j$). Hence, we are in one of the following two cases:

Case 1: We have $i = j + 1$.

Case 2: We have $i > j + 1$.

Let us first consider Case 1. In this case, $i = j + 1$. Thus, $w_i > w_{i+1}$ (which holds by assumption) rewrites as $w_{j+1} > w_{j+2}$. Moreover, the definition of j shows that $w_j > w_{j+1}$. Therefore, $w_j > w_{j+1} > w_{j+2}$. We shall use the notations $a := w_j$, $b := w_{j+1}$, $c := w_{j+2}$, $\ell := w_1 w_2 \cdots w_{j-1}$ and $r := w_{j+3} w_{j+4} \cdots w_n$; thus, $w = \ell abcr$ (since $w = w_1 w_2 \cdots w_n$) and $a > b > c$ (since $w_j > w_{j+1} > w_{j+2}$).

We have $w_1 w_2 \cdots w_{j-1} = \ell$ and $w_j = a$ and $w_{j+1} = b$ and $w_{j+2} w_{j+3} \cdots w_n = cr$ and therefore $w^{\frown j} = \ell bacr$. Thus, the equality (10) rewrites as

$$J(w) = \ell \langle a, b \rangle cr + J(\ell bacr). \tag{11}$$

Moreover, the word $\ell bacr$ is obtained from $\ell abcr = w$ by a single sorting step (since $a > b$); thus, (8) yields $\operatorname{inv}(\ell bacr) = \operatorname{inv} w - 1 < \operatorname{inv} w = k$. Hence, by the induction hypothesis, (9) holds for $\ell bacr$ instead of w. In particular, we can apply (9) to $\ell bacr$ and $j + 1$ instead of w and i (since the $(j + 1)$-st and $(j + 2)$-nd letters of $\ell bacr$ are a and c, and these satisfy $a > c$), and thus we obtain

$$J(\ell bacr) = \ell b \langle a, c \rangle r + J(\ell bcar). \tag{12}$$

Furthermore, the word $\ell bcar$ is obtained from $\ell bacr$ by a single sorting step (since $a > c$); thus, (8) yields $\operatorname{inv}(\ell bcar) = \operatorname{inv}(\ell bacr) - 1 < \operatorname{inv}(\ell bacr) < k$. Hence, by the induction hypothesis, (9) holds for $\ell bcar$ instead of w. In particular, we can apply (9) to $\ell bcar$ and j instead of w and i (since the j-th and $(j + 1)$-st letters of $\ell bcar$ are b and c, and these satisfy $b > c$), and thus we obtain

$$J(\ell bcar) = \ell \langle b, c \rangle ar + J(\ell cbar). \tag{13}$$

Now, (11) becomes

$$\begin{aligned} J(w) &= \ell \langle a, b \rangle cr + J(\ell bacr) = \ell \langle a, b \rangle cr + \ell b \langle a, c \rangle r + J(\ell bcar) \qquad \text{(by (12))} \\ &= \ell \langle a, b \rangle cr + \ell b \langle a, c \rangle r + \ell \langle b, c \rangle ar + J(\ell cbar) \qquad \text{(by (13))} \\ &= \ell \left(\langle a, b \rangle c + b \langle a, c \rangle + \langle b, c \rangle a \right) r + J(\ell cbar). \end{aligned} \tag{14}$$

On the other hand, our goal is to prove (9). In other words, our goal is to prove

$$J(w) = \ell a \langle b, c \rangle r + J(\ell acbr) \tag{15}$$

(since $i = j+1$, and thus $w_1 w_2 \cdots w_{i-1} = w_1 w_2 \cdots w_j = \ell a$ and $w_i = w_{j+1} = b$ and $w_{i+1} = w_{j+2} = c$ and $w_{i+2} w_{i+3} \cdots w_n = w_{j+3} w_{j+4} \cdots w_n = r$ and $w^{\frown i} = \ell acbr$). Let us rewrite the right hand side.

The word $\ell acbr$ is obtained from $\ell abcr = w$ by a single sorting step (since $b > c$); thus, (8) yields $\operatorname{inv}(\ell acbr) = \operatorname{inv} w - 1 < \operatorname{inv} w = k$. Hence, by the induction hypothesis, (9) holds for $\ell acbr$ instead of w. In particular, we can apply (9) to $\ell acbr$ and j instead of w and i (since the j-th and $(j + 1)$-st letters of $\ell acbr$ are a and c, and these satisfy $a > c$), and thus we obtain

$$J(\ell acbr) = \ell \langle a, c \rangle br + J(\ell cabr). \tag{16}$$

The word $\ell cabr$ is obtained from $\ell acbr$ by a single sorting step (since $a > c$); thus, (8) yields $\operatorname{inv}(\ell cabr) = \operatorname{inv}(\ell acbr) - 1 < \operatorname{inv}(\ell acbr) < k$. Hence, by the induction hypothesis, (9) holds for

$\ell cabr$ instead of w. In particular, we can apply (9) to $\ell cabr$ and $j+1$ instead of w and i (since the $(j+1)$-st and $(j+2)$-nd letters of $\ell cabr$ are a and b, and these satisfy $a > b$), and thus we obtain

$$J(\ell cabr) = \ell c \langle a, b\rangle r + J(\ell cbar).$$

Substituting this into (16), we find

$$J(\ell acbr) = \ell \langle a, c\rangle br + \ell c \langle a, b\rangle r + J(\ell cbar). \tag{17}$$

However, the axiom (L3) of the Lie structure A yields

$$(\langle a, b\rangle c - c\langle a, b\rangle) + (\langle b, c\rangle a - a\langle b, c\rangle) + (\langle c, a\rangle b - b\langle c, a\rangle) = 0.$$

In view of $\langle c, a\rangle = -\langle a, c\rangle$ (this is an easy consequence of axiom (L1), just as for Lie algebras), we can rewrite this as

$$(\langle a, b\rangle c - c\langle a, b\rangle) + (\langle b, c\rangle a - a\langle b, c\rangle) + ((-\langle a, c\rangle) b - b(-\langle a, c\rangle)) = 0.$$

Expanding the parentheses and collecting the negative terms on the right hand side, we transform this into

$$\langle a, b\rangle c + \langle b, c\rangle a + b\langle a, c\rangle = c\langle a, b\rangle + a\langle b, c\rangle + \langle a, c\rangle b. \tag{18}$$

Now, (14) becomes

$$\begin{aligned} J(w) &= \ell \underbrace{(\langle a, b\rangle c + b\langle a, c\rangle + \langle b, c\rangle a)}_{\substack{=\langle a,b\rangle c+\langle b,c\rangle a+b\langle a,c\rangle \\ =c\langle a,b\rangle+a\langle b,c\rangle+\langle a,c\rangle b \\ \text{(by (18))}}} r + J(\ell cbar) \\ &= \ell(c\langle a, b\rangle + a\langle b, c\rangle + \langle a, c\rangle b) r + J(\ell cbar) \\ &= \ell c\langle a, b\rangle r + \ell a\langle b, c\rangle r + \ell\langle a, c\rangle br + J(\ell cbar) \\ &= \ell a\langle b, c\rangle r + \underbrace{\ell\langle a, c\rangle br + \ell c\langle a, b\rangle r + J(\ell cbar)}_{\substack{=J(\ell acbr) \\ \text{(by (17))}}} \\ &= \ell a\langle b, c\rangle r + J(\ell acbr). \end{aligned}$$

This proves (15). In other words, (9) is proved for our arbitrary i (since (15) is just a rewritten form of (9)). This completes the induction step in Case 1.

Let us now consider Case 2. In this case, $i > j+1$. Thus, the four integers $j, j+1, i, i+1$ are distinct and ordered as follows: $j < j+1 < i < i+1$. We shall use the notations $a := w_j$, $b := w_{j+1}$, $c := w_i$, $d := w_{i+1}$, $\ell := w_1 w_2 \cdots w_{j-1}$, $m := w_{j+2} w_{j+3} \cdots w_{i-1}$ and $r := w_{i+2} w_{i+3} \cdots w_n$; thus, $w = \ell abmcdr$ (since $w = w_1 w_2 \cdots w_n$ and $j < j+1 < i < i+1$) and $a > b$ (since $w_j > w_{j+1}$) and $c > d$ (since $w_i > w_{i+1}$).

We have $w_1 w_2 \cdots w_{j-1} = \ell$ and $w_j = a$ and $w_{j+1} = b$ and $w_{j+2} w_{j+3} \cdots w_n = mcdr$ and thus $w^{\frown j} = \ell bamcdr$. Thus, the equality (10) rewrites as

$$J(w) = \ell\langle a, b\rangle mcdr + J(\ell bamcdr). \tag{19}$$

Moreover, the word $\ell bamcdr$ is obtained from $\ell abmcdr = w$ by a single sorting step (since $a > b$); thus, (8) yields $\operatorname{inv}(\ell bamcdr) = \operatorname{inv} w - 1 < \operatorname{inv} w = k$. Hence, by the induction hypothesis, (9) holds for $\ell bamcdr$ instead of w. In particular, we can apply (9) to $\ell bamcdr$ and i instead of w and

i (since the i-th and $(i+1)$-st letters of $\ell bamcdr$ are c and d, and these satisfy $c > d$), and thus we obtain

$$J(\ell bamcdr) = \ell bam \langle c, d\rangle r + J(\ell bamdcr).$$

Substituting this into (19), we find

$$\begin{aligned} J(w) &= \ell \langle a, b\rangle mcdr + \ell bam \langle c, d\rangle r + J(\ell bamdcr) \\ &= \ell (\langle a, b\rangle mcd + bam \langle c, d\rangle) r + J(\ell bamdcr). \end{aligned} \quad (20)$$

On the other hand, our goal is to prove (9). In other words, our goal is to prove

$$J(w) = \ell abm \langle c, d\rangle r + J(\ell abmdcr) \quad (21)$$

(since $w_1 w_2 \cdots w_{i-1} = \ell abm$ and $w_i = c$ and $w_{i+1} = d$ and $w_{i+2} w_{i+3} \cdots w_n = r$ and thus $w^{\frown i} = \ell abmdcr$). Let us rewrite the right hand side.

The word $\ell abmdcr$ is obtained from $\ell abmcdr = w$ by a single sorting step (since $c > d$); thus, (8) yields $\operatorname{inv}(\ell abmdcr) = \operatorname{inv} w - 1 < \operatorname{inv} w = k$. Hence, by the induction hypothesis, (9) holds for $\ell abmdcr$ instead of w. In particular, we can apply (9) to $\ell abmdcr$ and j instead of w and i (since the j-th and $(j+1)$-st letters of $\ell abmdcr$ are a and b, and these satisfy $a > b$), and thus we obtain

$$J(\ell abmdcr) = \ell \langle a, b\rangle mdcr + J(\ell bamdcr). \quad (22)$$

However, the axiom (L2) of the Lie structure A yields

$$\langle a, b\rangle m (cd - dc) = (ab - ba) m \langle c, d\rangle.$$

Expanding both sides, we rewrite this as

$$\langle a, b\rangle mcd - \langle a, b\rangle mdc = abm \langle c, d\rangle - bam \langle c, d\rangle.$$

Equivalently,

$$\langle a, b\rangle mcd + bam \langle c, d\rangle = abm \langle c, d\rangle + \langle a, b\rangle mdc.$$

Substituting this into (20), we obtain

$$\begin{aligned} J(w) &= \ell (abm \langle c, d\rangle + \langle a, b\rangle mdc) r + J(\ell bamdcr) \\ &= \ell abm \langle c, d\rangle r + \underbrace{\ell \langle a, b\rangle mdcr + J(\ell bamdcr)}_{\substack{= J(\ell abmdcr) \\ \text{(by (22))}}} \\ &= \ell abm \langle c, d\rangle r + J(\ell abmdcr). \end{aligned}$$

This proves (21). In other words, (9) is proved for our arbitrary i (since (21) is just a rewritten form of (9)). This completes the induction step in Case 2.

We have now completed the induction step in both Cases 1 and 2. Thus, the induction proof of (9) is complete. □

Thus, for each word $w = w_1 w_2 \cdots w_n$, we have defined an element $J(w) \in A$ that satisfies (9) for each $1 \leq i < n$ satisfying $w_i > w_{i+1}$ (not just for the smallest such i).

Next, let us define some general terminology. A word is said to be *sorted* if its letters are weakly increasing from left to right (i.e., if it has no inversions). For each word w, there is exactly one sorted word that contains the same letters (with the same multiplicities) as w; this word is called the *sorted rearrangement* of w, and can be obtained from w by successively applying single sorting steps until no more inversions remain. We shall use the notation $\operatorname{sort} w$ for this word. Note that

$$J(w) = 0 \qquad \text{for any sorted word } w \tag{23}$$

(by the base case of the recursive definition of $J(w)$, since $\operatorname{inv} w = 0$).

Furthermore, for any word w (or, more generally, any element w of $T(M)$), we shall let $\overline{w}$ denote the projection of w onto $S(M)$.

We shall now prove the following general formula: If v is a word and w is a sorted word, and $x \in X$ is a letter, then

$$(x \circ \overline{w})\, v = J(xwv) - J(\operatorname{sort}(xw)\, v). \tag{24}$$

Proof of (24). Indeed, let v be a word, let w be a sorted word, and let $x \in X$ be a letter. Note that $\overline{\operatorname{sort}(xw)} = \overline{xw}$ (since the words $\operatorname{sort}(xw)$ and xw differ only in the order of their letters, and thus have the same projection onto the commutative algebra $S(M)$).

Write the sorted word w as $w = t_1 t_2 \cdots t_n$. Thus, $\overline{w} = \overline{t_1 t_2 \cdots t_n} = \tau_1 \tau_2 \cdots \tau_n$, where each τ_i is the image of t_i in $S(M)$.

Since the word $t_1 t_2 \cdots t_n$ is sorted, we have $t_1 \leq t_2 \leq \cdots \leq t_n$. Thus, we have

$$t_1 \leq t_2 \leq \cdots \leq t_{m-1} < x \leq t_m \leq t_{m+1} \leq \cdots \leq t_n \tag{25}$$

for a unique integer $1 \leq m \leq n+1$ (in particular, $m = n+1$ if $t_n < x$, whereas $m = 1$ if $x \leq t_1$). Consider this m. Thus, $t_1 t_2 \cdots t_{m-1} x t_m t_{m+1} \cdots t_n$ is a sorted word (by (25)). The definition of $x \circ \overline{w}$ yields

$$\begin{aligned} x \circ \overline{w} &= \sum_{t_i < x} t_1 t_2 \cdots t_{i-1} \langle x, t_i \rangle t_{i+1} t_{i+2} \cdots t_n \qquad \left(\text{since } \overline{w} = \tau_1 \tau_2 \cdots \tau_n\right) \\ &= \sum_{i=1}^{m-1} t_1 t_2 \cdots t_{i-1} \langle x, t_i \rangle t_{i+1} t_{i+2} \cdots t_n \end{aligned} \tag{26}$$

(since the integers $i \in \{1, 2, \ldots, n\}$ satisfying $t_i < x$ are precisely $1, 2, \ldots, m-1$, according to (25)).

Moreover, from $w = t_1 t_2 \cdots t_n$, we obtain

$$\operatorname{sort}(xw) = \operatorname{sort}(x t_1 t_2 \cdots t_n) = t_1 t_2 \cdots t_{m-1} x t_m t_{m+1} \cdots t_n \tag{27}$$

(by (25)). Furthermore, the word xwv can be transformed into the word $\operatorname{sort}(xw)\, v$ by the following sequence of single sorting steps:

$$\begin{aligned} xwv &= x t_1 t_2 \cdots t_n v \qquad (\text{since } w = t_1 t_2 \cdots t_n) \\ &\mapsto t_1 x t_2 t_3 \cdots t_n v \qquad (\text{we swapped } x \text{ with } t_1) \\ &\mapsto t_1 t_2 x t_3 t_4 \cdots t_n v \qquad (\text{we swapped } x \text{ with } t_2) \\ &\mapsto \cdots \\ &\mapsto \underbrace{t_1 t_2 \cdots t_{m-1} x t_m t_{m+1} \cdots t_n}_{= \operatorname{sort}(xw)} v \qquad (\text{we swapped } x \text{ with } t_{m-1}) \\ &= \operatorname{sort}(xw)\, v. \end{aligned}$$

Using the recursive definition (9) of the J-function, we thus have

$$\begin{aligned} J\left(xt_1t_2\cdots t_nv\right) &= \left\langle x,t_1\right\rangle t_2t_3\cdots t_nv + J\left(t_1xt_2t_3\cdots t_nv\right);\\ J\left(t_1xt_2t_3\cdots t_nv\right) &= t_1\left\langle x,t_2\right\rangle t_3t_4\cdots t_nv + J\left(t_1t_2xt_3t_4\cdots t_nv\right);\\ &\ldots;\\ J\left(t_1t_2\cdots t_{m-2}xt_{m-1}t_m\cdots t_nv\right) &= t_1t_2\cdots t_{m-2}\left\langle x,t_{m-1}\right\rangle t_mt_{m+1}\cdots t_nv\\ &\qquad + J\left(t_1t_2\cdots t_{m-1}xt_mt_{m+1}\cdots t_nv\right). \end{aligned}$$

Substituting these equalities into one another, we obtain

$$\begin{aligned} &J\left(xt_1t_2\cdots t_nv\right)\\ &= \underbrace{\left\langle x,t_1\right\rangle t_2t_3\cdots t_nv + t_1\left\langle x,t_2\right\rangle t_3t_4\cdots t_nv + \cdots + t_1t_2\cdots t_{m-2}\left\langle x,t_{m-1}\right\rangle t_mt_{m+1}\cdots t_nv}_{=\sum_{i=1}^{m-1} t_1t_2\cdots t_{i-1}\left\langle x,t_i\right\rangle t_{i+1}t_{i+2}\cdots t_nv}\\ &\qquad + J\left(\underbrace{t_1t_2\cdots t_{m-1}xt_mt_{m+1}\cdots t_n}_{=\operatorname{sort}(xw)}v\right)\\ &= \underbrace{\sum_{i=1}^{m-1} t_1t_2\cdots t_{i-1}\left\langle x,t_i\right\rangle t_{i+1}t_{i+2}\cdots t_n}_{\substack{=x\circ\overline{w}\\ \text{(by (26))}}} v + J\left(\operatorname{sort}\left(xw\right)v\right)\\ &= \left(x\circ\overline{w}\right)v + J\left(\operatorname{sort}\left(xw\right)v\right). \end{aligned} \tag{28}$$

Since $t_1t_2\cdots t_n = w$, this rewrites as $J\left(xwv\right) = \left(x\circ\overline{w}\right)v + J\left(\operatorname{sort}\left(xw\right)v\right)$. This proves (24). $\square$

Similarly to (24), we can prove that if u is a word and w is a sorted word, and $y\in X$ is a letter, then

$$u\left(\overline{w}\circ y\right) = J\left(uwy\right) - J\left(u\operatorname{sort}\left(wy\right)\right). \tag{29}$$

Now, fix $x,y\in X$ and $x_1\leq x_2\leq\cdots\leq x_n$ in X. Let ξ, η and ξ_i be the images of x, y and x_i in $S\left(M\right)$. Let $\sigma = \xi_1\xi_2\cdots\xi_n$. We now take aim at proving the two equalities

$$\left(\xi\sigma\right)\circ y + \left(x\circ\sigma\right)y = J\left(xx_1x_2\cdots x_ny\right) \qquad \text{and} \tag{32}$$

$$x\circ\left(\sigma\eta\right) + x\left(\sigma\circ y\right) = J\left(xx_1x_2\cdots x_ny\right). \tag{33}$$

Once these two equalities are proved, it will immediately follow that

$$\left(\xi\sigma\right)\circ y + \left(x\circ\sigma\right)y = J\left(xx_1x_2\cdots x_ny\right) = x\circ\left(\sigma\eta\right) + x\left(\sigma\circ y\right),$$

thus

$$\left(\xi\sigma\right)\circ y - x\left(\sigma\circ y\right) = x\circ\left(\sigma\eta\right) - \left(x\circ\sigma\right)y,$$

and thus (3) will be proved. Thus, it remains to prove (32) and (33).

The word $x_1x_2\cdots x_n$ is sorted (since $x_1\leq x_2\leq\cdots\leq x_n$). Hence, applying (24) to $v=y$ and $w = x_1x_2\cdots x_n$, we find

$$\left(x\circ\overline{x_1x_2\cdots x_n}\right)y = J\left(xx_1x_2\cdots x_ny\right) - J\left(\operatorname{sort}\left(xx_1x_2\cdots x_n\right)y\right).$$

Thus,

$$\begin{aligned} J\left(xx_1x_2\cdots x_ny\right) &= J\left(\operatorname{sort}\left(xx_1x_2\cdots x_n\right)y\right) + \left(x\circ\overline{x_1x_2\cdots x_n}\right)y \\ &= J\left(\operatorname{sort}\left(xx_1x_2\cdots x_n\right)y\right) + \left(x\circ\sigma\right)y \end{aligned} \tag{34}$$

(since $\overline{x_1x_2\cdots x_n} = \xi_1\xi_2\cdots\xi_n = \sigma$). But $\operatorname{sort}\left(xx_1x_2\cdots x_n\right)$ is also a sorted word. Thus, applying (29) to $u = 1$ and $w = \operatorname{sort}\left(xx_1x_2\cdots x_n\right)$, we find

$$1\left(\overline{\operatorname{sort}\left(xx_1x_2\cdots x_n\right)}\circ y\right) = J\left(1\operatorname{sort}\left(xx_1x_2\cdots x_n\right)y\right) - J\left(1\operatorname{sort}\left(\operatorname{sort}\left(xx_1x_2\cdots x_n\right)y\right)\right).$$

This simplifies to

$$\begin{aligned} \overline{\operatorname{sort}\left(xx_1x_2\cdots x_n\right)}\circ y &= J\left(\operatorname{sort}\left(xx_1x_2\cdots x_n\right)y\right) - \underbrace{J\left(\operatorname{sort}\left(\operatorname{sort}\left(xx_1x_2\cdots x_n\right)y\right)\right)}_{\substack{=0 \\ \text{(by (23), since } \operatorname{sort}(\operatorname{sort}(xx_1x_2\cdots x_n)y) \\ \text{is a sorted word)}}} \\ &= J\left(\operatorname{sort}\left(xx_1x_2\cdots x_n\right)y\right). \end{aligned}$$

Thus,

$$J\left(\operatorname{sort}\left(xx_1x_2\cdots x_n\right)y\right) = \underbrace{\overline{\operatorname{sort}\left(xx_1x_2\cdots x_n\right)}}_{=\overline{xx_1x_2\cdots x_n}=\xi\xi_1\xi_2\cdots\xi_n=\xi\sigma}\circ y = (\xi\sigma)\circ y.$$

Substituting this into (34), we obtain

$$J\left(xx_1x_2\cdots x_ny\right) = (\xi\sigma)\circ y + (x\circ\sigma)\,y.$$

This proves (32). A similar argument (in which we first apply (29) to $u = x$ and $w = x_1x_2\cdots x_n$, and then apply (24) to $v = 1$ and $w = \operatorname{sort}\left(x_1x_2\cdots x_ny\right)$) proves (33). Hence, as we said above, (3) follows.

Appendix 2 (by DG). Birkhoff–Witt for flat modules

Theorem 7 (i) can be generalized as follows:

Theorem 10. *Let R be a fixed commutative ring, and M be a flat R-module. Then, $B(M) = 0$.*

Proof. The Govorov–Lazard theorem ([10, Théorème 1.2 (iii)], [9]) shows that M is a direct limit of a directed system $\{M_\alpha\}$ of free R-modules. By Theorem 7 (i), all these free R-modules M_α satisfy $B(M_\alpha) = 0$. Hence, Theorem 8 yields $B(M) = \varinjlim B(M_\alpha) = \varinjlim 0 = 0$. ☐

References

[1] Birkhoff, G. Representability of Lie algebras and Lie groups by matrices. *Annals Math.* **38** (1937), 526–532.

[2] Cartier, P. Remarques sur le théorème de Birkhoff–Witt. *Ann. Scuola norm. sup. Pisa, Sci. fis. mat.* 3 Ser. **12** (1958), 1–4.

[3] Cohn, P. M. A remark on the Birkhoff–Witt theorem. *J. London Math. Soc.* **38** (1963), 197–203.

[4] Fröhlich, A. Baer-invariants of algebras. *Trans. Am. Math. Soc.* **109** (1963), 221–244.

[5] Lazard, M. Sur les algèbres enveloppantes universelles de certaines algèbres de Lie. *Publ. Sci. Univ. Alger*, Sér. A. **1** (1954), 281–294.

[6] Nagata, M. "Local rings", (Interscience, New York, 1962).

[7] Witt, E. Treue Darstellung Liescher Ringe. *J. reine angew. Math.* **177** (1937), 152–160. [16]

Additional references inserted by DG:

[8] Baader, F. and Nipkow, T. "Term Rewriting and All That", (Cambridge University Press, Cambridge, 1998).

[9] Hines, R. Lazard's theorem (characterizing flatness), notes, December 7, 2016. `https://math.colorado.edu/~rohi1040/expository/lazardstheorem.pdf`

[10] Lazard, D. Autour de la platitude. *Bulletin de la Société Mathématique de France* **97** (1969), 81–128.

[11] Newman, M. H. A. On theories with a combinatorial definition of "equivalence". *Annals of Mathematics*, Second Series, **43** (1942), 223–243.

[16] *Comment by DG:* Some titles corrected.